



## Klasterisasi Data Ekonomi dan Demografi Menggunakan K-Means: Studi Kasus Saham, Neraca Pembayaran, GDP, dan Kependudukan

Dimas Rahma Samudra<sup>1</sup>, Jusuf Herlambang Herdiyana<sup>2</sup>, Muhammad Fauzan<sup>3</sup>, Aqiel Maulidan Zuhdi<sup>4</sup>,  
Muhammad Farraz Pradipta Bintang Kurniawan<sup>5</sup>, Zurnan Alfian<sup>6</sup>

<sup>1,2,3,4,5</sup> Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

e-mail : <sup>1</sup>dimassamudraa3344@gmail.com, <sup>2</sup>jusufherlamb@gmail.com, <sup>3</sup>mhmdfauzan86@gmail.com,

<sup>4</sup>maulidanzuhdi@gmail.com, <sup>5</sup>farrazpradipta@gmail.com, <sup>6</sup>dosen02678@unpam.ac.id

### Abstrak

Tujuan dari penelitian ini adalah untuk mengevaluasi seberapa baik dan fleksibel algoritma K-Means dalam mengidentifikasi pola tersembunyi pada berbagai data demografis dan ekonomi. Data seri waktu harga saham TLKM (2003), data makroekonomi neraca pembayaran AS (2000–2010), data PDB lintas sektoral negara-negara ASEAN (2014), riwayat kinerja perdagangan bulanan (2018), dan data populasi Kabupaten Minahasa Utara merupakan lima studi kasus yang dianalisis. Untuk memastikan kesetaraan bobot fitur, teknik studi mencakup langkah pra-pemrosesan dengan normalisasi Min-Max. Skor Silhouette kemudian digunakan untuk mengevaluasi secara objektif kualitas hasil pengelompokan. Hasil analisis menunjukkan bahwa K-Means secara konsisten menghasilkan kluster yang bermakna dan mudah dipahami di seluruh dataset. Algoritma ini dapat membedakan dua rezim pasar saham, membedakan antara periode booming ekonomi dan krisis, membagi negara-negara ASEAN menjadi tiga tingkat ekonomi, mengidentifikasi bulan-bulan dengan kinerja luar biasa, dan memetakan subdistrik berdasarkan profil demografis perkotaan dan pedesaan.

**Kata Kunci:** K-Means Clustering, Data Ekonomi, GDP ASEAN, Saham TLKM, Kependudukan

### 1. Pendahuluan

Era Digital ditandai dengan lonjakan volume data yang sebelumnya belum pernah terjadi di berbagai sektor, termasuk keuangan, ekonomi, dan sosial. Data yang beragam dari data mikro berfrekuensi tinggi seperti harga saham harian hingga data makro seperti Gross Domestic Product (GDP) dan indikator kependudukan, menyimpan berbagai pola yang sering tidak dapat diidentifikasi melalui analisis statik konvensional.

Pada konteks ini, *unsupervised learning* timbul berbagai paradigma penting dalam ilmu data, yang memungkinkan analisis pada data tanpa label atau kategori yang ditentukan sebelumnya. Salah satu dari teknik dalam *unsupervised learning* adalah analisis kluster, tujuan dari analisis kluster itu untuk mengelompokkan sekumpulan objek ke beberapa kluster. Tujuan utama dari analisis kluster itu untuk mempartisi data sedemikian rupa agar objek dalam satu kluster itu mempunyai tingkat kesamaan tinggi, sementara itu objek di kluster yang lain mempunyai tingkat kesamaan rendah. Proses ini membantu memahami struktur inheren dari data dan mengidentifikasi klasifikasi-klasifikasi tertentu.

Dari berbagai algoritma klasterisasi, *K-Means* adalah salah satu dari berbagai metode yang populer. Hal ini disebabkan pada konseptual yang sederhana, komputasi pada dataset besar yang efisien, dan implementasi yang mudah. Beberapa penelitian menunjukkan bahwa *K-Means* memiliki efektivitas dalam berbagai aplikasi, mulai dari klasifikasi pelanggan pada strategi pemasaran yang tepat, analisis ekonomi untuk mengelompokkan berbagai negara atau perusahaan berdasarkan finansialnya, dan untuk analisis data demografis untuk rancangan kebijakan publik.

Dari banyaknya penelitian yang telah menggunakan *K-Means*, rata-rata cenderung fokus pada tipe data tertentu. Kontribusi langsung dari penelitian ini adalah melakukan sistesis komparatif dari kinerja *K-Means* pada lima studi kasus dengan karakteristik data yang beragam. Kelima dataset tersebut dipilih untuk menguji kekuatan dan fleksibilitas algoritma, meliputi: (1) data time-series finansial berfrekuensi tinggi (Saham TLKM); (2) data time-series makroekonomi berfrekuensi rendah (Neraca Pembayaran AS); (3) data cross-sectional ekonomi (GDP ASEAN); (4) data time-series kinerja bulanan (Riwayat Trading); dan (5) data cross-sectional demografi (Kependudukan Minahasa Utara). Dengan membandingkan bagaimana *K-Means* mengidentifikasi pola struktur data yang berbeda dengan tujuan untuk mendapatkan pemahaman yang lebih dalam mengenai fleksibilitas algoritma.

Klasterisasi Data Ekonomi dan Demografi Menggunakan K-Means: Studi Kasus Saham, Neraca Pembayaran, GDP, dan Kependudukan

Berdasarkan latar belakang tersebut, rumusan masalah penelitian menjadi: “Seberapa efektif algoritma *K-Means* dalam mengidentifikasi berbagai kluster yang bermakna dan dapat diinterpretasikan dari lima dataset ekonomi dan demografi berbeda, dan wawasan apa yang dapat diambil dari pola tersebut untuk mendukung pengambilan keputusan di domain masing-masing?”. Penelitian ini diharapkan dapat memberikan bukti empiris mengenai kemampuan *K-Means* sebagai alat analisis data yang serbaguna.

**Rumusan masalah:** Bagaimana penerapan *K-Means* pada kelima dataset tersebut dan seberapa efektif algoritma ini dalam membentuk kluster yang bermakna untuk memetakan pola karakteristik data ekonomi dan demografi.

## 2. Metode Penelitian

Metodologi penelitian ini dirancang secara sistematis agar validitas dan konsistensi hasil analisis terjamin. Tahapan penelitian meliputi pengumpulan sumber data, *PreProcessing data*, penerapan algoritma *K-means*, menentukan jumlah kluster yang optimal, dan mengevaluasi kualitas kluster yang terbentuk.

### 2.1 Sumber Data

Penelitian ini menggunakan dataset nyata yang berasal dari sumber dan domain berbeda:

#### 2.1.1 Saham TLKM (2003)

Tabel 1. Dataset Saham TLKM

Tanggal	Open	High	Lowest	Close	%
7 April 2003	3.750	4.000	3.750	4.000	+6,7
8 April 2003	3.975	4.050	3.900	3.900	+2,5
9 April 2003	3.925	4.150	3.925	4.075	+4,5
10 April 2003	4.075	4.125	4.075	4.100	+0,6
11 April 2003	4.125	4.200	4.125	4.175	+4,3

Tabel data di atas menampilkan harga saham harian TLKM selama 5 hari, mulai dari 07 April 2003 sampai 11 April 2003. Data terdiri dari beberapa kolom. Open, High, Lowest, Close, dan % Persentase. Data diambil dari pasar saham (Bursa Efek Indonesia), berisi harga harian TLKM seperti Open, High, Low, dan Close. Dataset digunakan untuk mengamati perubahan harga (Close–Open) dan volatilitas (High–Low).

#### 2.1.2 Neraca Pembayaran Amerika Serikat (2000–2010)

Tabel 2. Dataset Neraca Pembayaran AS

Tahun	Goods Exports	Goods imports	Current Account	Capital Account	Total A to D
2000	775	-1227	-417	-1	0.31
2001	722	-1148	-385	-1	4.88
2002	666	-1167	-461	-1	3.71
2003	717	-1264	-523	-3	-1.33
2004	811	-1477	-625	-2	-2.80
2005	898	-1682	-729	-4	-14.10
2006	1020	-1863	-804	-4	-2.00
2007	1164	-1985	-718	0	0.00
2008	1312	-2139	-677	6	149.00
2009	1074	-1576	-377	0	84.00
2010	1293	-1936	-471	0	2.00

Bersumber dari buku *Multinational Business Finance Edisi ke-13* (data direkap manual). Fitur utama meliputi *Capital Account*, *Income Credit*, dan *Total A–D*. Data terdiri dari 11 observasi (tahun) dengan 12 variabel makroekonomi, meliputi:

- Transaksi Berjalan (Current Account),
- Akun Modal (Capital Account),
- Akun Keuangan (Financial Account),
- Total Neraca Pembayaran (Total A to D sebagai fitur utama klusterisasi).
-

### 2.1.3 GDP ASEAN (2014)

Tabel 3. Dataset GDP ASEAN

Negara	Jumlah Penduduk (Juta Jiwa)	GDP Per Kapita (US \$)
Singapura	6,90	56,286.8
Brunei Darussalam	0,40	41,703.0
Malaysia	30,89	10,933.5
Thailand	75,20	5,519.4
Indonesia	244,47	3,491.9
Filipina	102,80	2,870.5
Vietnam	90,39	2,052.3
Laos	6,38	1,446
Kamboja	15,25	934
Myanmar	60,67	835

Data diperoleh dari buku *Makro Ekonomi Pengantar untuk Manajemen* karya Detri Karya & Syamri Syamsuddin. Variabel mencakup jumlah penduduk (dalam juta jiwa) dan GDP (dalam USD).

### 2.1.4 Riwayat Trading Bulanan Tahun 2018 Dari myfxbook.com

Tabel 4. Dataset Riwayat Trading Bulanan

Bulan	Gain (%)	Kategori
Jan 2018	50.23%	Tinggi
Feb 2018	23.43%	Sedang
Mar 2018	25.2%	Sedang
Apr 2018	15.03%	Sedang
Mei 2018	24.51%	Sedang
Jun 2018	17.44%	Sedang
Jul 2018	17.73%	Sedang
Agu 2018	17.48%	Sedang
Sep 2018	2.55%	Rendah

Data ini diperoleh dari buku *Smart Way Forex Trading* Karya Alexander Hendi. Data pada tabel menunjukkan data keuntungan (gain) dalam bentuk persentase dari bulan Januari sampai September 2018, yang setelah itu dikelompokkan menjadi tiga kategori: tinggi, sedang, rendah.

### 2.1.5 Kondisi Penduduk Kabupaten Minahasa Utara

Tabel 5. Dataset Kondisi Penduduk Kabupaten Minahasa Utara

kecamatan	Penduduk Jiwa	Rumah Tangga	Luas Wilayah km <sup>2</sup>	Kepadatan Penduduk	Anggota Rumah Tangga
Kema	13499	3397	8257	163	4
Kauditan	22111	5768	9370	236	4
Airmadidi	24050	5986	7616	316	4
Kalawat	22381	5853	4901	457	4
Dimembe	21508	6255	12735	169	4
Talawaan	12556	3914	8098	155	3
Wori	18169	4768	8080	225	3
Likupan Barat	16952	4785	8743	194	4

Data diambil dari kantor Statistik Kota Manado Tahun 2008. Fitur yang digunakan: Jumlah penduduk, jumlah rumah tangga, luas kecamatan, kepadatan penduduk dan rata rata anggota per rumah tangga dari 9 kecamatan Kema, Kauditan, Armadidi, Kalawat, Dimembe, Talawaan, Wori, Likupan Barat dan Likupan Timur.

1. Fitur Utama:
  - Kepadatan per Km<sup>2</sup> → Langsung mencerminkan kepadatan penduduk.
2. Fitur Demografis Kunci:
  - Penduduk jiwa → Total populasi.
  - Rumah Tangga → Jumlah unit rumah tangga.
  - Anggota Rumah Tangga → Rata-rata anggota per rumah tangga .
3. Fitur Geografis Kritis:
  - Luas Wilayah km<sup>2</sup> → Area wilayah.

## 2.2 Preprocessing

Sebelum proses klusterisasi, data mentah harus melalui tahap pra-pemrosesan untuk memastikan kualitas dan kesiapannya untuk dianalisis. Dua langkah utama yang dilakukan adalah pemilihan fitur dan normalisasi data.

- **Pemilihan Fitur:** Dari semua dataset, hanya atribut numerik yang paling relevan dengan tujuan analisis yang dipilih. Pada beberapa kasus diperlukan rekayasa fitur untuk mendapatkan variabel yang lebih informatif.
- **Normalisasi Data:** Semua fitur numerik yang terpilih selanjutnya dinormalisasi menggunakan metode *Min-Max Normalization*. Proses ini mengubah skala setiap fitur ke rentang yang seragam antara 0 dan 1. Normalisasi ini merupakan langkah penting dalam algoritma *K-Means*. Tanpa normalisasi, fitur dengan skala nilai besar (misalnya, GDP perkapita dalam ribuan dolar) akan tidak proporsional mendominasi perhitungan jarak dibandingkan dengan skala nilai yang lebih kecil. Hal ini menyebabkan hasil klusterisasi menjadi bias. Dalam normalisasi, setiap fitur diberikan bobot yang sama ketika proses pengelompokannya, agar kluster yang terbentuk benar-benar didasarkan oleh kesamaan pola antar observasi.

## 2.3 Algoritma Klusterisasi

Prosedur dari *K-Means* dapat dijelaskan melalui beberapa langkah berikut:

- Menentukan jumlah kluster *K* yang diinginkan. Lalu, pilih *K* titik data secara acak dari dataset sebagai posisi *centroid* awal.
- Untuk setiap dataset, hitung jarak ke setiap *centroid*. Jarak yang paling biasa digunakan adalah jarak euclidean. Kemudian setiap titik data ditugaskan ke dalam kluster yang diwakili oleh *centroid* terdekat.
- Setelah titik data ditugaskan ke kluster, posisi *centroid* dihitung ulang. *Centroid* baru merupakan titik rata-rata dari seluruh data yang menjadi anggota klusternya.

## 2.4 Penentuan Jumlah Kluster Optimal

Tantangan utama dari menggunakan *K-Means* adalah menentukan jumlah kluster *K* di awal. Pada penelitian ini, *Elbow Method* digunakan untuk melakukan pendekatan heuristik untuk menemukan nilai *K* yang optimal. *Elbow Method* bekerja dengan cara menghitung metrik *Within-Cluster Sum of Squares* (WCSS) untuk berbagai nilai *K*. Pengertian dari WCSS itu sendiri adalah jumlah total dari kuadrat jarak antara setiap data dengan *centroid* di dalam klusternya.

## 2.5 Evaluasi Kualitas Kluster

Setelah semua kluster terbentuk, kualitas dari kluster itu harus dievaluasi secara kuantitatif. Penelitian ini menggunakan *Silhouette Score* sebagai metrik evaluasi utama. Metrik ini digunakan untuk mengukur seberapa akurat titik data ditempatkan dalam klusternya saat ini dibandingkan dengan kluster yang lain. Skor siluet untuk satu titik data dapat dihitung berdasarkan dua nilai: (a) jarak rata-rata ke semua titik lainnya dalam kluster yang sama, dan (b) jarak rata-rata ke semua titik dalam kluster lain yang terdekat. Interpretasi dari *Silhouette Score* adalah sebagai berikut:

- Skor mendekati +1 menunjukkan bahwa objek tersebut sangat cocok dengan klusternya sendiri dan sangat jauh dari kluster yang lain. Ini menunjukkan bahwa kluster itu padat dan terpisah dengan baik.
- Skor mendekati 0 menunjukkan bahwa objek itu berada di dekat antara batas dua kluster.
- Skor negatif menunjukkan bahwa objek itu mungkin telah salah diklasifikasikan dan lebih mirip dengan anggota kluster lain.

Nilai rata-rata *Silhouette Score* dari seluruh titik data dapat digunakan sebagai tolak ukur kualitas keseluruhan dari hasil klusterisasi untuk nilai *K* tertentu.

## 2.6 Tools

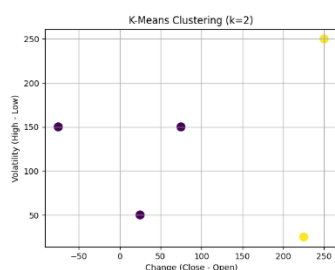
Semua proses analisis data, mulai dari *Preprocessing*, implementasi algoritma, hingga visualisasi hasil, menggunakan bahasa pemrograman *Python*. *Python* digunakan karena menyediakan berbagai *library* yang cukup lengkap. *Library* yang digunakan antara lain *pandas* untuk memanipulasi data, *scikit-learn* untuk implementasi *K-Means* dan metrik evaluasi, serta *matplotlib* dan *seaborn* untuk membuat visualisasi seperti plot *elbow* dan grafik kluster. Seluruh penelitian ini dijalankan menggunakan *Visual Studio Code*, sebagai editor utama.

## 3. Hasil Penelitian dan Pembahasan

Pada bagian ini disajikan hasil analisis klusterisasi yang telah dilakukan terhadap masing-masing dari kelima dataset yang digunakan dalam penelitian. Analisis ini difokuskan pada pola-pola yang berhasil diidentifikasi melalui proses pengelompokan data berdasarkan karakteristik tertentu. Setiap subbagian akan menguraikan jumlah kluster yang terbentuk, disertai interpretasi kualitatif terhadap masing-masing kluster guna memahami representasi data yang tersembunyi di dalamnya.

Selain itu, pembahasan akan dilengkapi dengan analisis mendalam mengenai makna dari tiap kluster, relevansinya terhadap tujuan penelitian, serta implikasi yang mungkin timbul baik dari segi praktis maupun teoritis. Temuan-temuan ini diharapkan dapat memberikan wawasan yang lebih luas dalam memahami struktur data serta mendukung pengambilan keputusan yang lebih tepat sasaran berdasarkan hasil klusterisasi tersebut.

### 3.1 Analisis Klusterisasi Data Saham TLKM (2003)



Gambar 1 Visualisasi *K-Means Clustering* Data Saham TLKM

Analisis pertama fokus pada data pergerakan saham harian PT Telkom Indonesia Tbk (TLKM) selama tahun 2003. Tujuan dari analisis ini adalah mengidentifikasi pola “tipe hari” perdagangan yang berulang berdasarkan perubahan harga (Close-Open) dan volatilitas (High-Low).

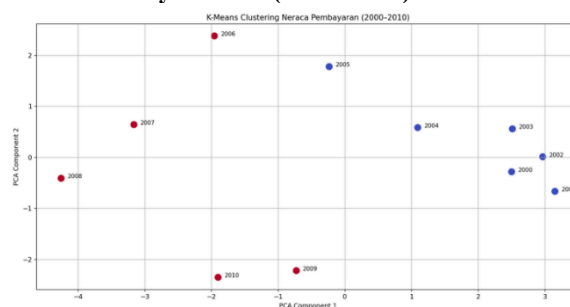
- **Hasil Kuantitatif:** Berdasarkan analisis, jumlah kluster yang digunakan pada dataset ini adalah  $k=2$ . Visualisasi menunjukkan dua kelompok pola perdagangan yang lumayan berbeda, yang memungkinkan indentifikasi rezim pasar yang berbeda.
- **Interpretasi Kluster:** Karakteristik dari kluster disajikan pada gambar 1, yang menjadi dasar interpretasi kualitatif.

#### Pembahasan:

- **Klaster 0 (Hari Normal/Konsolidasi):** Pada kluster ini (ditandai dengan warna ungu) mencakup mayoritas hari perdagangan. Pola ini mewakili kondisi pasar yang relatif normal, dilihat dari pergerakan harga yang normal atau tidak terlalu ekstrem. Ini merupakan hari konsolidasi dengan pergerakan moderat. Kluster ini menunjukkan tidak adanya sentimen pasar yang kuat hanya ke satu arah.
- **Klaster 1 (Hari Tren Bullish Ekstrem):** Kluster ini (ditandai dengan warna kuning) menandakan hari-hari yang luar biasa. Ciri utamanya adalah dengan kenaikan harga (Close-Open) yang cukup signifikan. Kenaikan ekstrem ini terjadi dalam dua skenario volatilitas: sangat rendah (kenaikan yang stabil setiap hari) atau sangat tinggi (kenaikan yang disertai lonjakan harga yang besar). Hari-hari ini menggambarkan sentimen *bullish* yang sangat kuat yang kemungkinan dipicu oleh berita atau peristiwa pasar yang signifikan.

**Implikasi:** Kedua klasifikasi tersebut membedakan antara kondisi pasar “biasa” dan “luar biasa”. Bagi analisis, hal ini dapat digunakan untuk membangun model yang berbeda. Strategi yang dirancang untuk kluster 0 mungkin tidak efektif jika dilakukan untuk kluster 1, dan begitupun sebaliknya. Pengenalan dua kluster ini memnungkinkan pengembangan strategi perdagangan yang lebih efektif.

### 3.2 Analisis Klasterisasi Neraca Pembayaran AS (2000–2010)



Gambar 2 Visualisasi *K-Means Clustering* Neraca Pembayaran AS

Analisis yang kedua ini menerapkan *K-Means* pada data makro ekonomi tahunan Neraca Pembayaran Amerika Serikat dari tahun 2000 hingga 2010. Dengan tujuan untuk melihat apakah algoritma dapat mengidentifikasi periode ekonomi yang berbeda.

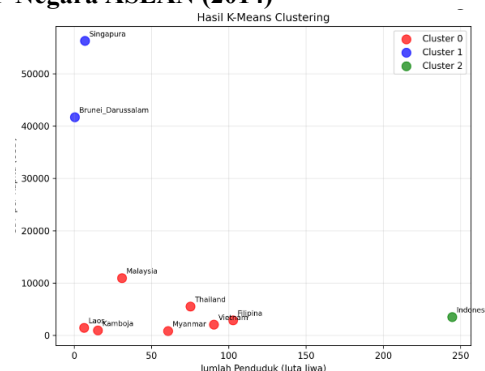
- **Hasil Kuantitatif:** Analisis ini dilakukan dengan  $K=2$ . Visual yang dihasilkan dari reduksi dimensi (PCA) menampilkan dua kluster yang jelas dan terpisah, yang menandakan adanya dua periode ekonomi yang secara struktur sangat berbeda dalam rentang waktu tersebut.
- **Interpretasi Kluster:** gambar 2 menunjukkan tahun dari kedua kluster.

#### Pembahasan:

- **Kluster 0 (Fase Ekspansi Ekonomi):** Pada kluster ini (ditandai dengan warna biru) yang mencakup tahun-tahun dari 2000 sampai dengan 2005. Periode ini dikenal dengan sebagai masa ekspansi ekonomi, yang ditandai dengan defisit neraca berjalan yang dialiri modal masuk yang kuat. Titik-titik data dalam kluster ini berkelompok secara rapat, menampilkan stabilitas ekonomi selama periode tersebut.
- **Kluster 1 (Fase Gejolak dan Krisis):** Pada kluster ini (ditandai dengan warna merah) dengan tepat menangkap periode mengarah ke, selama, dan setelah krisis finansial global 2008. Dimulai pada tahun 2006, kluster ini memperlihatkan pergeseran struktural dalam ekonomi Amerika Serikat. Pada tahun 2006 dan 2007 dapat dilihat sebagai periode “*build-up*” menuju krisis, sedangkan 2008 merupakan puncak dari krisis, dan 2009-2010 merupakan fase awal pemulihan. Persebaran titik yang lebih luas dalam kluster ini secara visual menggambarkan volatilitas dan ketidakpastian ekonomi yang meningkat lebih tinggi selama periode ini.

Algoritma *K-Means*, berhasil mengidentifikasi pergeseran periode ekonomi. Titik pemisah antara kedua kluster secara tepat menandai mulainya periode ketidakstabilan yang akhirnya berujung pada krisis finansial.

### 3.3 Analisis Klasterisasi GDP Negara ASEAN (2014)



Gambar 3 Visualisasi *K-Means Clustering* GDP ASEAN 2014

Analisis yang ketiga fokus kepada data *cross-sectional* yang membandingkan negara ASEAN pada tahun 2014 berdasarkan GDP per kapita dan jumlah penduduk. Tujuan dari analisis ini adalah mengelompokkan negara dengan profil ekonomi yang serupa.

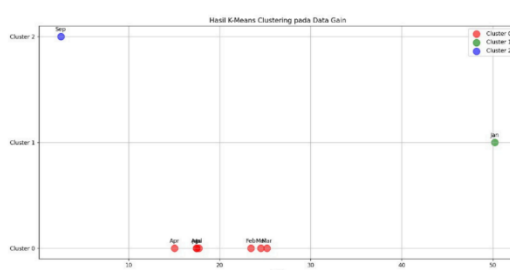
- **Hasil Kuantitatif:** Analisis ini dilakukan dengan  $K=3$ . Hasil visual ini menggambarkan tiga kluster yang sangat jelas, mengindikasikan bahwa pengelompokan negara ASEAN dalam tiga tingkatan ekonomi yang sangat valid dan didukung kuat oleh data.
- **Interpretasi Kluster:** Gambar 3 menunjukkan karakteristik dari ketiga kluster ekonomi ASEAN

#### Pembahasan:

- **Klaster 1 (Ekonomi Berpenghasilan Tinggi):** Pada klaster ini (biru) secara jelas menunjukkan Singapura dan Brunei, yang menonjol karena GDP per kapita yang sangat tinggi dengan populasi yang kecil.
- **Klaster 0 (Ekonomi Berkembang & Menengah):** Pada klaster ini (merah) adalah kelompok terbesar, mencakup negara dengan GDP per kapita yang lebih rendah. Kelompok ini dalam hal tingkat pendapatan per kapita yang relatif lebih rendah dibandingkan dengan klaster lainnya.
- **Klaster 2 (Outlier Populasi Tinggi):** Klaster ini (hijau) mengidentifikasi Indonesia sebagai klaster tersendiri. GDP per kapita Indonesia berada diantara negara dengan klaster 0, namun dengan jumlah penduduk yang banyak (lebih dari 240 juta jiwa) menjadikannya sebuah *outlier* yang signifikan secara statistik. Algoritma ini menyatakan bahwa profil demografi-ekonomi Indonesia unik dan tidak bisa disamakan dengan negara ASEAN lainnya.

**Implikasi:** Indonesia teridentifikasi sebagai klasifikasi yang unik dan sangat penting untuk perusahaan yang mempertimbangkan skala pasar. Klaster 1 adalah pasar untuk produk premium, klaster 2 (Indonesia) adalah pasar massal yang sangat masif, dan klaster 0 adalah kelompok pasar berkembang dengan karakteristik yang sangat beragam.

### 3.4 Analisis Klasterisasi Riwayat Trading Bulanan (2018) Dari myfxbook.com



Gambar 4 Visualisasi *K-Means Clustering* Riwayat Trading Bulanan 2018

Analisis yang keempat menerapkan *K-Means* pada dataset keuntungan (*gain*) bulanan dari sebuah riwayat *trading* selama tahun 2018 untuk dikategorikan bulan-bulan berdasarkan tingkat kinerjanya..

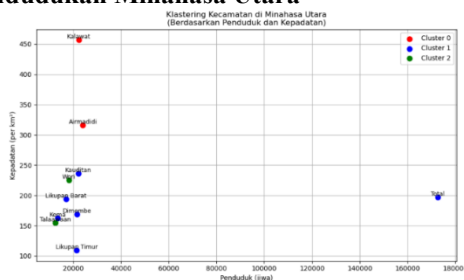
- **Hasil Kuantitatif:** Analisis ini dilakukan dengan  $K=3$ . Hasil visual jelas memisahkan kinerja bulanan dalam tiga kelompok berbeda di sepanjang sumbu “*Gain*”.
- **Interpretasi Klaster:** Pada gambar 4 ini adalah karakteristik dari ketiga klaster kinerja.

**Pembahasan:** Hasil klasterisasi secara efektif mengidentifikasi tiga periode kinerja yang berbeda:

- **Klaster 1 (Kinerja Tinggi):** Pada klaster ini (hijau) berisi hanya bulan Januari, yang merupakan *outlier* positif dengan keuntungan yang sangat tinggi.
- **Klaster 2 (Kinerja Rendah):** Pada klaster ini (biru) berisi hanya bulan September, merupakan *outlier* negatif dengan kinerja yang paling rendah dalam setahun.
- **Klaster 0 (Kinerja Sedang):** Pada klaster ini (merah) mencakup mayoritas bulan dalam setahun, mewakili periode kinerja yang “normal” atau stabil.

**Implikasi:** Analisis ini menunjukkan keahlian *K-Means* untuk mengisolasi *outlier* secara efektif. Hal ini dapat mempermudah trader untuk membuat strategi spesifik untuk melakukan pekerjaannya.

### 3.5 Analisis Klasterisasi Kependudukan Minahasa Utara



Gambar 5 Visualisasi *K-Means Clustering* Data Kependudukan Minahasa Utara

Analisis yang terakhir ini menerapkan *K-Means* pada data demografi tingkat kecamatan di Kabupaten Minahasa Utara untuk melakukan klasifikasi geografis

- **Hasil Kuantitatif:** Analisis ini dilakukan dengan  $K=3$ . Hasil visual menunjukkan tiga kelompok kecamatan dengan profil demokrasi yang berbeda berdasarkan kepadatan dan jumlah penduduk.
- **nterpretasi Klaster:** gambar ke 5 ini menunjukkan karakteristik dari ketiga klasifikasi demografis

**Pembahasan:**

- **Klaster 0 (Pusat Urban Padat):** Pada klaster ini (merah) dengan sangat jelas menggambarkan Kalawat dan Airmadidi sebagai pusat kota kabupaten. Dengan ciri khas yang kepadatan penduduknya jauh lebih tinggi dibandingkan dengan yang lain, meskipun dengan jumlah penduduk absolut yang tidak besar.
- **Klaster 1 (Wilayah Suburban & Rural):** Pada klaster ini (biru) merupakan kelompok terbesar yang mencakup mayoritas kecamatan. Ciri-cirinya adalah kepadatan penduduk yang lebih rendah dan tersebar di wilayah yang lumayan luas. Klaster ini bisa dianggap sebagai wilayah penyangga di sekitar pusat perkotaan.
- **Klaster 2 (Rural Populasi Rendah):** Klaster ini (hijau) mengisolasi Kecamatan Kema sebagai klaster tersendiri. Meskipun kepadatannya tidak serendah beberapa daerah pada klaster 1, jumlah penduduknya yang paling kecil menjadikannya sebuah *outlier* demografis.

**Implikasi:** Klasifikasi ini menjelaskan gambaran demografis yang lebih mendalam bagi pengembangan daerah. Pemerintah daerah dapat mengidentifikasi tiga kategori wilayah yang berbeda: (1) Pusat kota yang ramai yang memerlukan fokus pada infrastruktur perkotaan (klaster 0), (2) Area pedesaan yang luas dan beragam yang memerlukan strategi pembangunan yang lebih umum (klaster 1), dan (3) Satu daerah dengan karakteristik khusus dan populasi terbatas (klaster 2) yang memerlukan penanganan atau program yang dirancang khusus.

**3.6 Sintesis dan Evaluasi Komparatif**

Dataset	Tipe Data	Optimal	Interpretasi Kualitas Klaster
Saham TLKM	Time-series, Frekuensi Tinggi	2	Cukup (Dua rezim pasar teridentifikasi)
Neraca Pembayaran AS	Time-series, Frekuensi Rendah	2	Sangat Baik (Pemisahan rezim ekonomi sangat jelas)
GDP ASEAN	Cross-sectional	3	Sangat Baik (Struktur hierarki dan outlier sangat jelas)
Riwayat Trading	Time-series, Frekuensi Rendah	3	Baik (Outlier kinerja terisolasi dengan baik)
Kependudukan Minahasa Utara	Cross-sectional	3	Cukup Baik (Klasifikasi urban dan outlier teridentifikasi)

Analisis komparatif ini memberikan informasi yang lebih mendalam terkait kinerja *K-Means*. Algoritma ini menggambarkan performa terbaik pada data yang memiliki kelompok yang jelas dan terpisah secara alami, contohnya pada data GDP ASEAN dan Neraca Pembayaran Amerika Serikat. Kemampuannya untuk mengisolasi *outlier* sebagai klaster tersendiri salah satu keahlian utamanya dalam analisis data.

Pada data yang di mana batas antara kelompok yang tidak terlalu tajam, hasilnya mungkin tidak sejelas itu, namun tetap menghasilkan ilmu yang berharga. Hal ini menandai bahwa berhasilnya *K-Means* tidak hanya bergantung pada algoritma itu sendiri, tapi juga pada struktur yang melekat di dalam data.

**4. Kesimpulan**

Penelitian ini telah berhasil menerapkan algoritma klasterisasi K-Means pada lima dataset dengan karakteristik yang beragam, mencakup data finansial, ekonomi makro, performa trading, dan demografi. Hasil analisis menunjukkan bahwa K-Means mampu mengidentifikasi pola tersembunyi yang relevan dan dapat diinterpretasikan dengan baik pada seluruh studi kasus. Secara khusus, algoritma ini berhasil mengelompokkan perilaku pasar saham TLKM menjadi dua rezim yang berbeda, mengungkap dua fase ekonomi Amerika Serikat yang dipisahkan oleh krisis finansial 2008, membentuk tiga klaster ekonomi negara-negara ASEAN berdasarkan GDP dan populasi, mengkategorikan performa trading bulanan ke dalam tiga kelompok kinerja, serta memetakan wilayah-wilayah demografis di Minahasa Utara berdasarkan kepadatan dan jumlah penduduk. Kemampuan K-Means dalam menemukan struktur data yang bermakna mencerminkan fleksibilitas dan efektivitasnya sebagai alat analisis data yang kuat. Implikasi praktis dari temuan ini cukup luas, mulai dari penyusunan strategi pasar berdasarkan klasifikasi tren harga saham, peringatan dini terhadap perubahan siklus ekonomi, hingga perencanaan tata kota dan pembangunan daerah berbasis pemetaan demografis. Analisis berbasis klasterisasi seperti ini

memungkinkan pengambilan keputusan yang lebih terarah dan berbasis data di berbagai sektor. Namun demikian, penelitian ini juga menyadari adanya beberapa keterbatasan. Kinerja K-Means sangat bergantung pada pemilihan pusat awal dan jumlah kluster yang ditentukan sejak awal. Selain itu, asumsi bentuk kluster yang cenderung bulat dan ukuran yang seragam membuat algoritma ini kurang efektif dalam kondisi data yang memiliki distribusi tidak merata atau mengandung noise.

#### Daftar Pustaka

- Febby Rachmawati, Y. N. (2024). Buletin Ilmiah Mat. Stat. dan Terapannya. *ANALISIS CLUSTER MENGGUNAKAN METODE K-MEANS DAN FUZZY C-MEANS*, 629 – 638.
- Maycandra Desdianti, N. N. (2024). Buletin Ilmiah Math. Stat. dan Terapannya. *ANALISIS K-MEANS CLUSTERING DENGAN METODE ELBOW PADA PENGELOMPOKAN TINGKAT PENGANGGURAN DI KALIMANTAN BARAT*, 53-62.
- Eiteman, D. K., Stonehill, A. I., & Moffett, M. H. (2016). *MULTINATIONAL BUSINESS FINANCE* (13TH EDITION). Pearson. 120.
- Hasmanto, B. (2008). *Implementasi Regulasi Administrasi Kependudukan*. Tangerang: CV. Cipta utama Tangerang.
- Hendi, A. (2019). *Smart Way Forex Trading*. Jakarta: PT ELEX MEDIA KOMPUTINDO.
- Karya, D., & Syamsuddin I, S. (2016). *Makro Ekonomi Pengantar untuk manajemen*. Depok: PT. RajaGrafindo Persada.
- Tambunan, A. P. (2007). *Menilai Harga Wajar Saham (Stock Valuation)*. Jakarta: PT Alex Media Komputindo.